蜜蜂球囊菌的参考转录组 *de novo* 组装及 SSR 分子标记开发

张曌楠[1,#], 熊翠玲[1,#], 徐细建[1], 黄枳腱[1], 郑燕珍[1], 骆　群[2], 刘　敏[1], 李汶东[1], 童新宇[1], 张　琦[1], 梁　勤[1], 郭　睿[1,*], 陈大福[1,*]

(1. 福建农林大学蜂学学院, 福州 350002; 2. 江西省养蜂研究所, 南昌 330201)

摘要:【目的】通过 RNA seq 技术对纯培养的蜜蜂球囊菌 *Ascosphaera apis* 孢子和球囊菌侵染的蜜蜂幼虫肠道组织进行测序, *de novo* 组装球囊菌的参考转录组, 并对其进行功能与代谢通路注释, 进而基于该转录组数据开发球囊菌的 SSR 分子标记。【方法】首先通过差速离心获得活化的球囊菌孢子, 配制含 1×10^7 孢子/mL 的饲料饲喂 4, 5 和 6 日龄的意大利蜜蜂 *Apis mellifera ligustica* 幼虫和中华蜜蜂 *Apis cerana cerana* 幼虫, 通过 Illumina HiSeq™ 2500 平台同时对上述蜜蜂幼虫肠道及纯化球囊菌孢子进行深度测序, 原始数据过滤后通过 Trinity 软件组装得到 unigenes, 进而通过 BLASTX 比对 NCBI Nr, Swiss-Prot, KOG 和 KEGG 数据库对 unigenes 进行功能与代谢通路注释。利用 MISA 软件对所有 unigenes 进行 SSR 搜索, 并利用 Primer Premier 5 软件设计 SSR 特异性引物, 通过 PCR 对不同来源的球囊菌 SSR 位点进行扩增。【结果】本研究共获得 146 135 308 条高质量 reads, *de novo* 组装得到 42 609 个 unigenes。BLASTX 比对结果显示, 29 316 个 unigenes 在上述公共数据库中具有功能和代谢通路注释。注释到法夫酵母 *Xanthophyllomyces dendrorhous* 上的 unigenes 最多, 达 6 050个。KEGG 注释结果显示, unigenes 可注释到 117 个代谢通路, 其中富集在核糖体(ribosome)上的 unigenes 数量最多(529)。所有 unigenes 中共预测到 7 968 个 SSRs, 通过 PCR 开发出 5 个球囊菌 SSR 分子标记。【结论】本研究成功组装球囊菌的参考转录组, 并进行了功能与代谢通路注释, 可为在分子水平深入研究球囊菌提供重要的参考信息。基于该转录组信息开发的 5 个 SSR 分子标记可推动菌株鉴定、基因图谱构建及基因定位等研究。

关键词: 蜜蜂球囊菌; 参考转录组; RNA seq; 功能与代谢通路注释; SSR

中图分类号: Q966　　**文献标识码:** A　　**文章编号:** 0454-6296(2017)01-0034-11

De novo assembly of a reference transcriptome and development of SSR markers for *Ascosphaera apis*

ZHANG Zhao-Nan[1,#], XIONG Cui-Ling[1,#], XU Xi-Jian[1], HUANG Zhi-Jian[1], ZHENG Yan-Zhen[1], LUO Qun[2], LIU Min[1], LI Wen-Dong[1], TONG Xin-Yu[1], ZHANG Qi[1], LIANG Qin[1], GUO Rui[1,*], CHEN Da-Fu[1,*] (1. College of Bee Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. Apiculture Institute of Jiangxi Province, Nanchang 330201, China)

Abstract:【Aim】This study aims to deeply sequence the purified spores of *Ascosphaera apis* and honeybee larval gut infected with *A. apis* using RNA seq technology, to *de novo* assemble a reference transcriptome of *A. apis*, to further conduct functional and metabolic pathway annotation, and finally to develop SSR markers for *A. apis*.【Methods】*A. apis* spores purified after differential centrifugation were

---

基金项目: 现代农业产业技术体系建设专项资金(CARS-45-KXJ7); 福建农林大学科技发展资金(KF2015123); 福建省教育厅科技项目(JA10104); 福建省大学生创新创业训练计划项目(201610389053)

作者简介: 张曌楠, 男, 1991 年 7 月生, 河南新乡人, 硕士研究生, 主要从事蜜蜂保护学研究, E-mail: 15236055367@163.com; 熊翠玲, 女, 1977 年 1 月生, 福建永定人, 实验师, 主要从事蜜蜂保护学研究, E-mail: pandacl@sina.com

#共同第一作者 Authors with equal contribution

* 通讯作者 Corresponding authors, E-mail: fafu_ruiguo@126.com; dfchen826@163.com

收稿日期 Received: 2016-11-02; 接受日期 Accepted: 2016-12-13

fed to the 4-, 5- and 6-day-old larvae of *Apis mellifera ligustica* and *A. cerana cerana* at a concentration of 1×10^7 spore/mL. The purified spores of *A. apis* and the infected honeybee larval guts were sequenced at Illumina HiSeq™ 2500 platform. After filtration of raw reads, unigenes were *de novo* assembled with Trinity software and annotated through BLASTX against the NCBI Nr, Swiss-Prot, KOG and KEGG databases. The MISA software was used to search SSRs in all unigenes, and Primer Premier 5 software was used to design specific primers for SSR loci. PCR was carried out to amplify SSR loci in *A. apis* from different regions in China. 【Results】In this study, a total of 146 135 308 high-quality reads were produced from RNA seq, and 42 609 unigenes were *de novo* assembled. BLASTX results indicated that 29 316 unigenes have function or metabolic pathway annotations in the aforementioned public databases. And the most unigenes (6 050) were annotated to *Xanthophyllomyces dendrorhous*. The results of KEGG pathway annotation revealed that all unigenes were annotated to 117 metabolic pathways, among which ribosome (529 unigenes) was the most enriched one. In total, 7 968 SSRs were predicted from all unigenes, and five SSR markers were developed by PCR amplification. 【Conclusion】In this study, a reference transcriptome of *A. apis* was successfully *de novo* assembled and annotated, providing significant information for further studying *A. apis* at the molecular level. Five SSRs developed in this study can facilitate the research on *A. apis* including strain identification, gene map construction and gene localization.

Key words: *Ascosphaera apis*; reference transcriptome; RNA seq; function and metabolic pathway annotation; SSR

蜜蜂白垩病(chalkbrood disease)是一种最具代表性的蜜蜂真菌病,1913 年 Massen 在德国首次报道白垩病,中国大陆 1990 年发生白垩病(梁勤和陈大福, 2009),近年来,随着蜂产品贸易的快速发展,白垩病呈逐年上升的趋势(Aizen *et al*., 2009)。该病是由蜜蜂球囊菌 *Ascosphaera apis* 特异性侵染蜜蜂幼虫而导致,发病时期一般为春季和初夏,虽然不对蜂群造成毁灭性打击,但却能造成成年蜜蜂数量的大幅下降,从而严重影响蜂蜜等产品的产量(Bailey and Ball, 1991; Wood, 1998)。据报道,白垩病可造成蜂蜜产量下降 5% ~37% (Zaghloul *et al*., 2005)。近几十年,人们从形态学(Spiltoir, 1955; Skou, 1988; Anderson and Gibson, 1998)、培养方法(Anderson and Gibson, 1998; Laf, 2015)、流行病学(Flores *et al*., 2005a, 2005b)、增殖方式(Pöggeler, 2001; Chorbiński, 2003)、发病机理(Bailey and Ball, 1991; Theantana and Chantawannakul, 2008)、免疫机制(Stanley *et al*., 2009; Evans and Spivak, 2010)及防治策略(Hornitzky, 2001; Tarpy, 2002; Mourad *et al*., 2005)等方面对白垩病进行了较多研究并取得一些重要进展。然而,目前尚无一种抗真菌药物被批准用于养蜂生产,一些替代的防治方法,如饲养管理(Anderson *et al*., 1997; Baggio *et al*., 2005)等,虽取得一定效果,但总体并不理想。

本课题组也对球囊菌病原检测等方面进行了一系列研究,如李江红等(2012)给人工饲养的意大利蜜蜂 *Apis mellifera ligustica*(简称意蜂)幼虫接种球囊菌孢子,发现 28℃低温处理能够显著提高处于幼虫到蛹转化期蜜蜂的发病率,摄入的孢子在整个幼虫期因中肠没有氧气而不生长,幼虫的取食和发育过程正常,至幼虫期结束进入蛹期后,蜜蜂的中后肠接通,摄入的孢子伴随蜜蜂的蛹便进入后肠并在此迅速萌发生长,1 ~2 d 内菌丝即突破体表而导致蜜蜂死亡;郑志阳等(2011)对健康和患病蜜蜂幼虫血淋巴进行 SDS-PAGE 电泳和蛋白酶、酯酶的活性染色,发现健康蜜蜂幼虫血淋巴中的蛋白含量丰富,主要由 4 种高分子质量的蛋白组成,而患病幼虫血淋巴中的蛋白含量很少,主要蛋白组分被降解。由于基因组信息的缺失,球囊菌在分子水平的研究近乎停滞。Qin 等(2006)公布了球囊菌 ARSEF7406 菌株的基因组序列,为球囊菌的分子研究奠定了基础,但作者并没有公布基因的位置和功能注释信息。Cornman 等(2012)利用 Roche 454 焦磷酸测序技术对来自培养基的球囊菌菌丝和来自蜜蜂幼虫感染组织的球囊菌菌丝进行了转录组测序,功能分析表明球囊菌的差异表达基因参与了交配类型、细胞内信号转导和应激反应。

鉴于目前球囊菌参考基因组位置及功能注释信息缺失的现实,本研究利用 RNA seq 技术对纯培养的球囊菌孢子进行深度测序,同时对球囊菌胁迫的

意蜂幼虫及中华蜜蜂 *Apis cerana cerana*（简称中蜂）幼虫肠道组织进行测序，通过比对意蜂的参考基因组和中蜂幼虫肠道的参考转录组过滤掉蜜蜂幼虫肠道组织的测序数据，将过滤后的测序数据与球囊菌孢子的测序数据混合，进而 *de novo* 组装球囊菌的 unigenes，并对其进行功能注释与代谢通路注释。利用本研究组装得到的球囊菌参考转录组，可对纯培养的球囊菌与感染肠道内的球囊菌进行差异表达基因分析，进而通过转录组水平的深入分析揭示球囊菌胁迫过程中的病原-宿主互作。本研究将为球囊菌的分子研究提供重要的参考信息，基于该参考转录组信息开发出的 5 个 SSR 分子标记，可用于球囊菌菌株鉴定、基因图谱构建及遗传多样性等研究。

1 材料与方法

1.1 供试蜜蜂与球囊菌

本研究使用的意蜂幼虫和中蜂幼虫取自福建农林大学蜂学学院教学蜂场，蜜蜂球囊菌菌株由福建农林大学蜂学学院蜜蜂保护实验室保存并活化。

1.2 主要实验试剂及仪器

DEPC 水购自生工®生物工程（上海）有限公司，DNaseI 和 Oligotex mRNA Kits Midi 试剂盒购自德国 Qiagen 公司，Dynal M280 磁珠购自美国 Invitrogen 公司，高碘酸钠购自美国 Sigma 公司，DNA Ligase 购自美国 Thermo 公司，RNAiso Reagent 抽提试剂盒、Ex Taq Polymerase 及 Superscript Ⅱ Reverse Transcriptase 均购自日本 TaKaRa 公司，纯化 cDNA 的 Agencourt AMPure XPbeads 购自 Beckman 美国公司，cDNA 文库构建试剂盒 TruSeq™ DNA Sample Prep Kit v2-Set A 购自美国 Illumina 公司，DNA marker 为日本 TaKaRa 公司产品。其他试剂均为国产分析纯。

恒温恒湿气候箱购自中国宁波江南仪器厂，pH 酸度计购自中国上海仪电科学股份有限公司，超纯水仪购自中国四川沃特尔水处理设备有限公司，高速冷冻离心机购自德国 Eppendorf 公司，倒置显微镜购自中国上海光学仪器五厂，超净工作台购自中国苏州安泰空气技术有限公司，PCR 仪购自美国 Bio Rad 公司，凝胶成像仪购自中国上海培清科技有限公司，超低温冰箱购自中国中科美菱低温科技股份有限公司。

1.3 蜜蜂幼虫饲养及球囊菌侵染

意蜂幼虫及中蜂幼虫饲料配方参照王倩等（2009）的研究报道。7 日龄两种蜜蜂幼虫的成活率可达 70% 和 90% 以上。将 2 日龄幼虫移至 24 孔板培养板，置于恒温恒湿培养箱，35℃，70% 相对湿度（RH）条件下饲养，每 24 h 更换饲料。3 日龄时，一组幼虫饲喂含球囊菌孢子的人工饲料（孢子浓度为 1×10^7 孢子/mL），另一组饲喂正常人工饲料。上述两组均设 3 个生物学重复。

1.4 测序样品的准备

从接菌活化的平板上刮取球囊菌孢子至 EP 管中，按照 Jensen 等（2013）的方法进行纯化球囊菌孢子，液氮速冻后迅速转移至超低温冰箱保存备用。分别剖取上述球囊菌侵染的意蜂幼虫（4，5 和 6 日龄）肠道与中蜂幼虫（4，5 和 6 日龄）肠道，液氮速冻后迅速转移至超低温冰箱保存备用。

1.5 cDNA 文库构建及 RNA seq 测序

首先提取上述各球囊菌侵染肠道样品与球囊菌纯化孢子样品总 RNA，用带有 Oligo（dT）的磁珠富集真核生物 mRNA。加入 fragmentation buffer 将 mRNA 打断成短片段，以 mRNA 为模板，用六碱基随机引物（random hexamers）合成 cDNA 第 1 链，然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成 cDNA 第 2 链，在经过 QiaQuick PCR 试剂盒纯化并加 EB 缓冲液洗脱之后做末端修复、加 poly（A）并连接测序接头，然后用琼脂糖凝胶电泳进行片段大小选择，最后进行 PCR 扩增，建好的测序文库用 Illumina HiSeq™ 2500 进行测序。

1.6 球囊菌参考转录组的 *de novo* 组装

对于球囊菌侵染意蜂、中蜂幼虫肠道样品的测序数据，首先分别比对意蜂参考基因组（assembly Amel_4.5，assembly ACSNU-2.0）和中蜂幼虫肠道参考转录组（本课题组组装，原始数据已上传 NCBI SRA 数据库，SRA Num：SRA456721），将未比对上的数据与纯化球囊菌孢子测序数据混合。利用 Perl 脚本去除上述混合数据中含有 adaptor、N 的比例大于 10% 的和低质量读段（reads），获得有效读段（clean reads）。利用软件 Trinity 进行球囊菌 unigenes 的 *de novo* 组装（缺省值 Kmer = 25）。长度短语 200 bp 的 contigs 和 unigene 将被舍弃。过滤和组装以后得到高质量的 unigene。RNA seq 原始数据已上传 NCBI SRA 数据库，SRA Num：SRA464366。

1.7 Unigenes 基本功能注释

利用 BLASTX（E-value $<10^{-5}$）将测序序列比对 NCBI Nr，Swiss-Prot，KOG 和 KEGG 数据库。通过 BLASTX 将组装出来的 unigenes 序列与 Nr 数据库

进行比对后，取每个 unigene 在 Nr 库中比对结果最好（E 值最低）的那一条序列为对应同源序列（如有并列，取第一条）确定同源序列所属物种，统计比对到各个物种的同源序列数量。基于 Nr 数据库注释结果，利用 Blast2GO 进行 unigenes 的 GO 注释，利用 WEGO 软件进行 GO 分类。

通过 BLASTX（E-value $< 10^{-5}$ 程序将 unigenes 序列比对至 Nr，KOG，KEGG 和 Swiss-Prot 蛋白数据库，最佳比对结果将用于确定 unigenes 的方向。如果不同库之间的比对结果有矛盾，则按 Nr，Swiss-Prot，KEGG 和 KOG 的优先级确定 unigenes 的序列方向。若某一转录本无法比对到上述数据库，利用 ESTscan 预测 unigenes 的编码区及确定序列方向。

1.8　SSR 分子标记开发

利用软件 MISA 搜索 unigenes 的微卫星标记，按照以下标准从 unigenes 中查找 SSR 位点：二核苷酸重复≥6 次，三核苷酸重复≥5 次，四核苷酸重复≥5 次，五核苷酸重复≥5 次和六核苷酸重复≥5 次，同时分析 SSR 的基元类型的特征。根据 MISA 的输出结果，利用 Primer Premier 5（Premier 公司，加拿大）对每一个含有 16 bp 碱基重复的 SSR 设计引物。

选取来源于福建福州（FZ）和浙江苍山（CS）的球囊菌样本进行 SSR 位点鉴定，首先利用真菌 RNA 抽提试剂盒 RNAiso Reagent（TaKaRa 公司，日本）抽提球囊菌孢子总 RNA，然后利用反转录试剂盒 Superscript Ⅱ Reverse Transcriptase（TaKaRa 公司，日本）进行反转录得到 cDNA，随机选择 20 对 SSR 特异性引物进行 PCR 扩增。PCR 反应体系为 20 μL：cDNA 模板 1 μL，Mixture 10 μL，上下游引物（10 μmol/L）各 1 μL，无菌水补至 20 μL；PCR 程序：94℃预变性 5 min；94℃变性 50 s，55℃退火 30 s，72℃延伸 30 s，共 33 个循环；最后 72℃延伸 10 min。PCR 产物经 1% 琼脂糖凝胶电泳检测。

2　结果

2.1　RNA seq 数据分析及 unigenes 组装

球囊菌纯化孢子与球囊菌侵染蜜蜂幼虫肠道组织 Illumina 测序，共获得 151 312 584 条原始读段（raw reads），去除含 adapter 的、含 N 比例大于 10% 的和低质量的 reads，共得到 146 135 308 条 clean reads，各样品 clean reads 数均在 27 141 980（94.37%）以上，两端 Q20（99% 碱基正确率）均在 3 292 721 137（97.05%）以上，说明 RNA seq 数据质量良好（表 1）。

将 clean reads 组装为重叠群（contigs）进一步对 contigs 序列拼接和去冗余处理分别得到 42 609 个 unigenes（表 2），平均长度为 966 nt，N50 为 1 550 nt（图 1），说明本研究 unigenes 的组装质量较好。经统计，unigene 的数目随着序列长度的增加而减少，在 300 ~ 399 nt 长度范围内数目最多，符合生物体序列长度分布的基本规律。

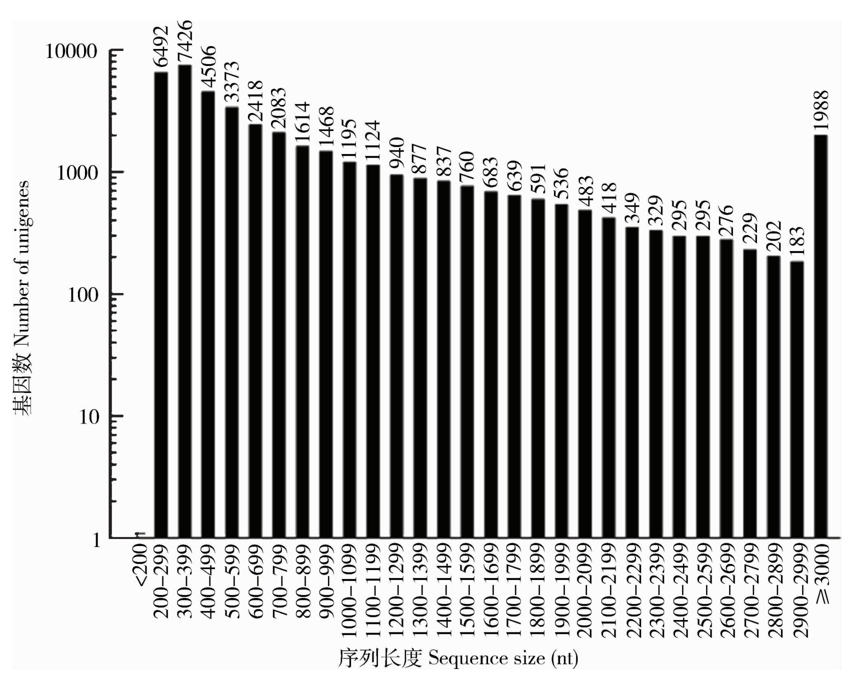


图 1　球囊菌的 unigenes 长度分布图

Fig. 1　Length distribution of unigenes of *Ascosphaera apis*

表 1 RNA seq 数据统计

Table 1 Summary of RNA seq data in this research

样本 Sample	过滤前 reads 数目 Number of reads before filtration	过滤后 reads 数目(百分比) Number of reads after filtration (percentage)	GC 百分比 GC percentage	接头数(百分比) Number of adapters (percentage)	低质量 reads 数(百分比) Number of low-quality reads (percentage)	Q20 百分比 Q20 percentage
AaCK-1	27 346 616	258 80 356 (94.64%)	52.87	121 010 (0.44%)	1 341 932 (4.91%)	97.11
AaCK-2	28 762 462	27 141 980 (94.37%)	55.38	142 288 (0.50%)	1 474 808 (5.13%)	97.05
AaCK-3	28 615 498	27 069 220 (94.6%)	55.41	124 140 (0.44%)	1 418 572 (4.96%)	97.15
AamT1-1	2 528 840	2 469 622 (97.66%)	41.94	6 528 (0.26%)	52 690 (2.08%)	97.45
AamT1-2	2 737 168	2 693 106 (98.39%)	43.02	5 668 (0.21%)	38 394 (1.40%)	97.85
AamT1-3	3 023 132	2 980 240 (98.58%)	43.25	5 532 (0.18%)	37 360 (1.24%)	98.18
AamT2-1	3 178 228	3 122 140 (98.24%)	42.28	6 568 (0.20%)	49 520 (1.56%)	97.99
AamT2-2	2 547 204	2 501 890 (98.22%)	43.29	5 928 (0.23%)	39 386 (1.55%)	97.75
AamT2-3	2 814 364	2 768 034 (98.35%)	42.33	5 510 (0.20%)	40 820 (1.45%)	98.07
AamT3-1	3 229 530	3 178 978 (98.43%)	44.91	6 086 (0.19%)	44 466 (1.38%)	97.89
AamT3-2	2 345 336	2 253 412 (96.08%)	42.95	7 566 (0.33%)	84 358 (3.60%)	97.08
AamT3-3	3 050 274	2 942 398 (96.46%)	43.12	7 862 (0.26%)	100 014 (3.28%)	97.19
AacT1-1	2 799 036	2 799 036 (100%)	42.56	0 (0%)	0 (0%)	97.14
AacT1-2	3 196 972	3 196 972 (100%)	42.02	0 (0%)	0 (0%)	97.29
AacT1-3	3 189 982	3 189 982 (100%)	41.85	0 (0%)	0 (0%)	97.08
AacT2-1	4 146 604	4 146 604 (100%)	43.28	0 (0%)	0 (0%)	98.01
AacT2-2	3 424 686	3 424 686 (100%)	44.86	0 (0%)	0 (0%)	97.72
AacT2-3	3 653 650	3 653 650 (100%)	45.28	0 (0%)	0 (0%)	97.91
AacT3-1	9 455 236	9 455 236 (100%)	49.38	0 (0%)	0 (0%)	98.02
AacT3-2	4 398 766	4 398 766 (100%)	45.82	0 (0%)	0 (0%)	97.70
AacT3-3	6 869 000	6 869 000 (100%)	47.18	0 (0%)	0 (0%)	97.95

表 2 Unigenes 组装结果统计表

Table 2 Summary of *de novo* assembly of unigenes

基因数目 Gene number	GC 百分比 GC percentage	N50 数值 N50	最大长度(nt) Maximum length	最小长度(nt) Minimum length	平均长度(nt) Average length	总碱基数 Total number of assembled bases
42 609	50.0160	1 550	12 740	179	966	41 192 931

2.2 Unigenes 注释

利用 BLASTX(E-value < 10^{-5})将 unigenes 序列比对 NCBI Nr, Swiss-Prot, KOG 和 KEGG 数据库,比对结果显示分别有 24 027, 24 311, 20 931 和 12 983 个 unigenes 能够注释到上述公共蛋白数据库,有功能注释的 unigenes 数目为 29 316,占全部 unigenes 的 68.8%;此外,有 31.19% 的 unigenes 无功能注释(表 3)。

表 3 四大蛋白数据库中 unigenes 的注释统计

Table 3 Summary of annotation information of all unigenes in four public protein database

所有通用基因 Total unigenes	Nr	Swiss-Prot	KOG	KEGG	注释基因 Annotated genes	未注释基因 Unannotated genes
42 610	24 027	24 311	20 931	12 983	29 316	13 294

注释到 Nr 数据库中 unigenes 的 E-value 分布显示(图 2),比对到物种序列的 E-value 均小于 1E-5,其中 E-value 小于 1E-100 的有 43.27%,说明注释结果可信度较高。注释基因同源序列的物种分布统计结果显示,注释基因数最多的物种是法夫酵母 *Xanthophyllomyces dendrorhous*,其次是白僵菌 *Beauveria bassiana* D1-5 和伞枝犁头霉 *Lichtheimia corymbifera* JMRC:FSU:9682)(表 4)。

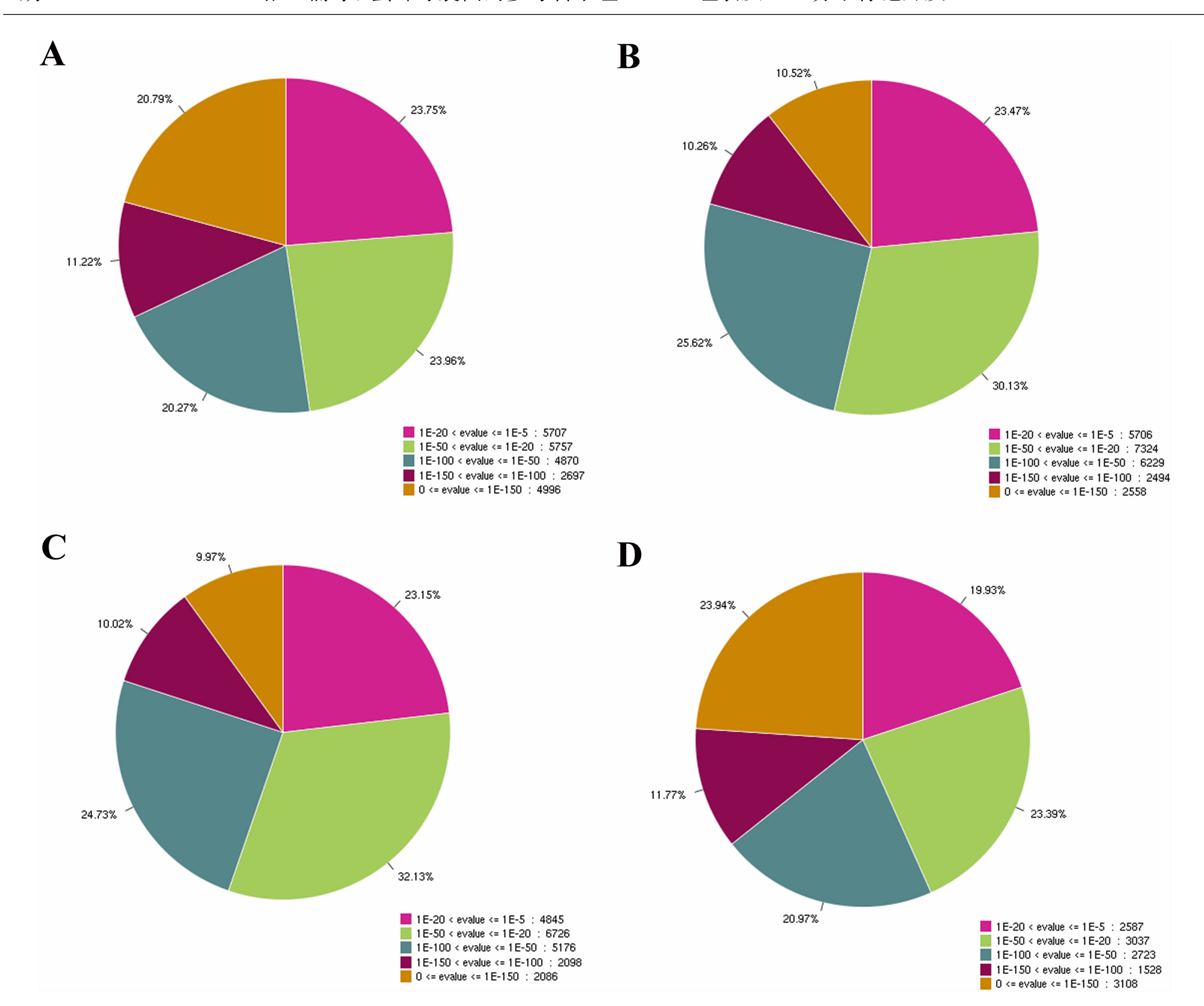


图 2　Nr（A），Swiss-Prot（B），KOG（C）和 KEGG（D）数据库中 unigenes 的 E 值分布

Fig. 2　Distribution of E-value of unigenes in databases Nr（A），Swiss-Prot（B），KOG（C） and KEGG（D）

表 4　Unigenes 在各物种分布统计表

Table 4　Distribution of unigenes in different species

物种 Species	基因数目 Number of unigenes
法夫酵母 *Xanthophyllomyces dendrorhous*	6 050
球孢白僵菌 *Beauveria bassiana* D1-5	745
伞枝犁头霉 *Lichtheimia corymbifera* JMRC:FSU:9682	649
毛霉菌 *Mucor ambiguus*	542
根瘤菌 *Rhizophagus irregularis* DAOM 197198w	535
球孢子菌 *Coccidioides immitis* RS	422
兹壶菌 *Rozella allomycis* CSF55	403
副球孢子菌 *Paracoccidioides* sp. 'lutzii' Pb01	397
皮炎芽生菌 *Blastomyces dermatitidis* SLH14081	385
青霉菌 *Penicillium rubens* Wisconsin 54-1255	330

2.3　Unigenes 的 Gene Ontology（GO）分类

利用 WEGO 软件对可注释到 GO 数据库的 unigenes 进行分类，结果显示这些 unigenes 主要分为生物学进程、细胞组分和分子组分 3 个大类，共同富集在 40 个 GO term 上（图 3），基因富集数最多的前十个 GO term 分别为细胞进程（cellular process）、代谢过程（metabolic process）、催化活动（catalytic activity）、细胞整体（cell）、细胞部分（cell part）、结合活性（binding activity）、单有机体过程（single-organism process）、细胞器（organelle）、定位（localization）以及大分子复合物（macromolecular complex）（表 5）。

2.4　Unigenes 的 KEGG 代谢通路注释

对所有 unigenes 进行 KEGG 代谢通路富集分析，发现共有 7 694 个 unigenes 注释到 KEGG 数据库中（图 4）。这些 unigenes 分布于 117 个已知的代谢通路中，其中基因富集数最多的 10 个代谢通路分别为核糖体（ribosome）、氨基酸生物合成

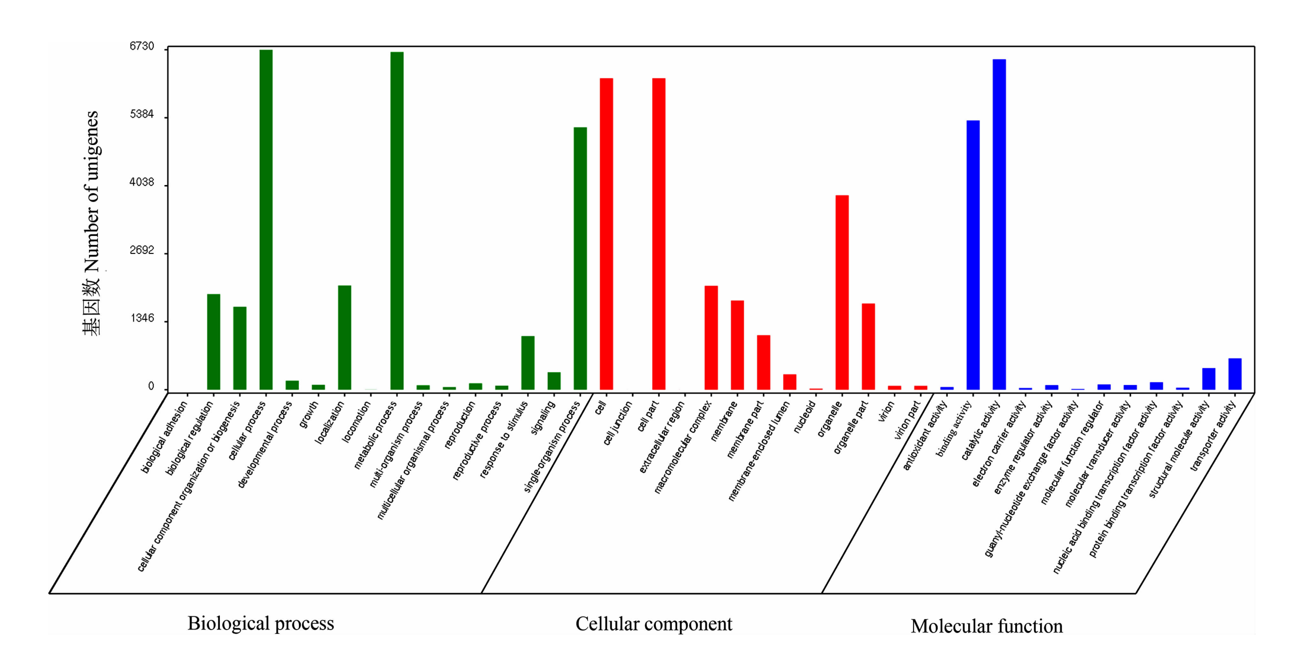


图 3 全部 unigenes 的 GO 分类

Fig. 3 GO classification of all unigenes

KEGG pathway annotation

Pathway	Number
新陈代谢 Metabolism	
碳水化合物代谢 Carbohydrate metabolism	974
新陈代谢总览 Global and overview	798
氨基酸代谢 Amino acid metabolism	789
能量代谢 Energy metabolism	530
脂代谢 Lipid metabolism	489
核苷酸代谢 Nucleotide metabolism	427
辅因子与维生素代谢 Metabolism of cofactors and vitamins	402
其他氨基酸代谢 Metabolism of other amino acids	245
多糖生物合成与代谢 Glycan biosynthesis and metabolism	237
萜类化合物与聚酮化合物代谢 Terpenoids and polyketides metabolism	77
其他次生代谢产物生物合成 Biosynthesis of other secondary metabolites	37
遗传信息加工 Genetic information processing	
翻译 Translation	1343
折叠、分类与降解 Folding, sorting and degradation	1060
转录 Transcription	553
复制与修复 Replication and repair	322
环境信息加工 Environmental information processing	
信号转导 Signal transduction	196
膜转运 Membrane transport	43
细胞进程 Cellular processes	
转运与代谢 Transport and metabolism	883
细胞生长与死亡 Cell growth and death	319
有机体系统 Organismal systems	
衰老 Aging	15
人类疾病 Human diseases	
内分泌与代谢疾病 Endoctrine and metabolic diseases	109
抗药性 Drug resistance	2

0 269 538 807 1076 1345

图 4 全部 unigenes 的 KEGG 代谢通路注释结果

Fig. 4 KEGG metabolic pathway annotation of all unigenes

表 5　基因富集数前 10 位的 GO term

Table 5　Top 10 GO classification of unigenes

分类 Class	数目 Number
细胞进程 Cellular process	6 729
新陈代谢 Metabolic process	6 686
催化活动 Catalytic activity	6 543
细胞整体 Cell	6 167
细胞部分 Cell part	6 167
结合活性 Binding activity	5 332
单有机体过程 Single-organism process	5 195
细胞器 Organelle	3 850
定位 Localization	2 065
大分子复合物 Macromolecular complex	2 057

(biosynthesis of amino acids)、碳代谢(carbon metabolism)、内质网蛋白加工(protein processing in endoplasmic reticulum)、剪接体(spliceosome)、RNA 转运(RNA transport)、嘌呤代谢(purine metabolism)、内吞(endocytosis)、氧化磷酸化(oxidative phosphorylation)和泛素介导的蛋白水解(ubiquitin mediated proteolysis)(表 6)。

2.5　SSR 分子标记开发

利用 MISA 软件对所有 unigenes 进行搜索,共预测出 7 968 个 SSR 位点,其中三核苷酸(tri-nucleotide)重复最多,数目达到 4 235,其次是二核苷酸(di-nucleotide, 2 575)(表 7)。SSR 基元分析结果显示,AG/CT 的出现频率最高,其次是 AT/AT 和 AGC/CTG(图 5)。

表 6　基因富集数最多的前 10 个 KEGG 代谢通路

Table 6　Top 10 metabolic pathway enrichment of unigenes in KEGG database

代谢通路 Metabolic pathway	基因富集数 All genes with pathway annotation	代谢通路号 Pathway ID
核糖体 Ribosome	529	ko03010
氨基酸生物合成 Biosynthesis of amino acids	459	ko01230
碳代谢 Carbon metabolism	452	ko01200
内质网蛋白质加工 Protein processing in endoplasmic reticulum	421	ko04141
剪接体 Spliceosome	407	ko03040
RNA 转运 RNA transport	388	ko03013
嘌呤代谢 Purine metabolism	379	ko00230
内吞 Endocytosis	351	ko04144
氧化磷酸化 Oxidative phosphorylation	343	ko00190
泛素介导蛋白水解 Ubiquitin mediated proteolysis	283	ko04120

表 7　球囊菌 SSR 位点统计

Table 7　Statistics of SSRs of *Ascosphaera apis*

统计项目 Statistics item	数目 Number
检测序列总数 Total number of sequences examined	42 610
检测序列总长度(bp)Total size of examined sequences	41 192 931
鉴定出的 SSR 总数 Total number of identified SSRs	7 968
包含 SSR 的序列总数 Number of SSR containing sequences	5 233
含 1 个以上 SSR 的序列总数 Number of sequences containing more than one SSR	1 615
化合物形成过程中的 SSR 总数 Number of SSRs present in compound formation	1 012
二核苷酸 Di-nucleotide	2 575
三核苷酸 Tri-nucleotide	4 235
四核苷酸 Tetra-nucleotide	674
五核苷酸 Penta-nucleotide	264
六核苷酸 Hexa-nucleotide	220

随机选择 20 对特异性引物,对来源于福州和苍山的球囊菌样本进行 PCR 扩增。PCR 产物经 1% 琼脂糖凝胶电泳检测,结果显示,共有 5 对引物(表 8)可在不同来源球囊菌样品中二者都扩增出符合预期的特异性片段(图 6),说明这些鉴定出的 SSR 位点可作为球囊菌特有的分子标记。

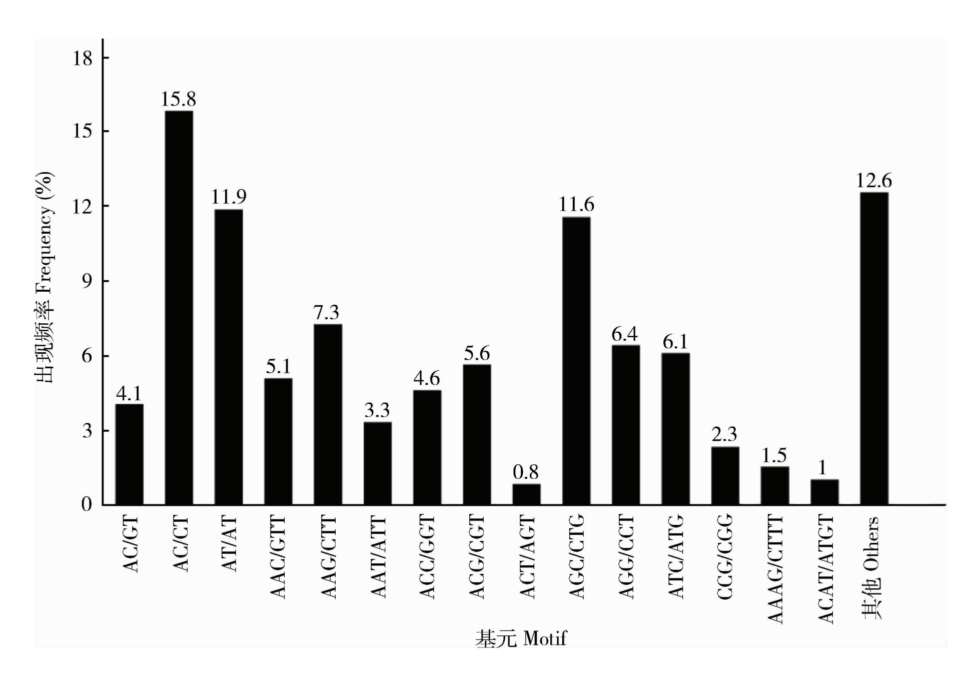


图 5 不同串联重复单元类型的 SSR 在总 SSR 中所占比例统计

Fig. 5 Different types of tandem repeat units of SSRs in all SSRs

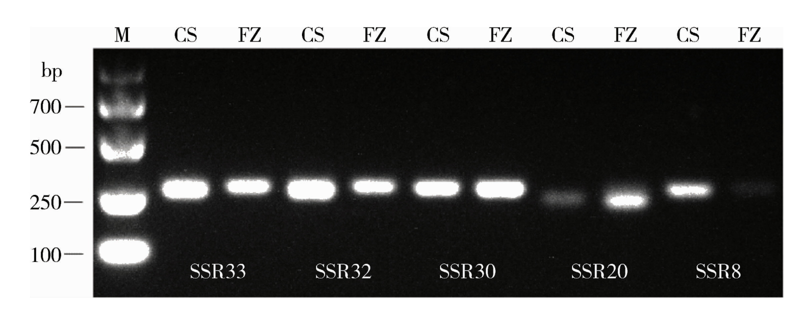


图 6 不同来源球囊菌 SSR 位点鉴定

Fig. 6 Identification of SSRs loci in *Ascosphaera apis* samples from two different regions in China

CS：浙江苍山样品 Samples from Cangshan，Zhejiang；FZ：福建福州样品 Samples from Fuzhou，Fujian.

表 8 SSRs 引物序列

Table 8 Primers used in SSR locus amplification

引物名称 Primer name	引物序列(5′-3′) Primer sequence	预期片段大小(bp) Expected fragment size
Uni8-F	CCTCCAATGTACCATCACCC-	239
Uni8-R	TGAGGGTTGTTGATGCGATA	
Uni20-F	GCTTGAAGCGTAATCCTTCG	215
Uni20-R	CAGGCTGCACTGGAGTATCA	
Uni30-F	TCCTCGATTTCGTAACCGTC	252
Uni30-R	TGTTCCAGAAATTGGGAAGG	
Uni32-F	TGCTGATGCTGATGAAGAGG	254
Uni32-R	GCAAGTGGCTGACATGAAGA	
Uni33-F	CGGTTACCGCGATCAATTAC	271
Uni33-R	AGCCTCGAACGAACGAATTA	

3 讨论

本课题组前期已对球囊菌胁迫的意蜂幼虫和中蜂幼虫进行转录组研究(未发表数据)，为了揭示球囊菌胁迫过程中的病原-宿主互作，需要对球囊菌进行转录组学研究。虽然球囊菌的基因组早在 2006 年就已公布(Qin *et al.*, 2006)，但当时并未公布基因位置及功能注释信息，导致球囊菌基因功能研究进展缓慢。本研究利用 RNA seq 技术对纯培养的球囊菌和球囊菌侵染的蜜蜂幼虫肠道进行深度测序，之所以同时对后者进行测序，目的是尽可能增大 RNA seq数据量，从而提升球囊菌参考转录组的组

装质量。本研究共组装得到 42 609 个 unigenes，平均长度为 966 nt，N50 达到 1 550 nt，较好的测序质量可为球囊菌的转录组学研究提供可靠的参考信息，也可以用于验证球囊菌基因组信息。通过比对 NCBI Nr，Swiss-Prot，KOG 数据库对组装的球囊菌参考转录组进行功能注释，通过比对 KEGG 数据库对其进行代谢通路注释，有功能及代谢通路注释的 unigenes 数目为 29 316（68.8%），这些注释将为后续在分子水平研究球囊菌提供重要的参考信息。此外，有 13 294（31.19%）个 unigenes 没有功能注释，填补这些空白有赖于上述公共数据库信息的不断丰富和完善。本研究中，球囊菌注释基因数最多的物种是法夫酵母，后者的分子研究起步较早且较为深入（Baeza *et al.*，2012；Alcaíno *et al.*，2014；Gutiérrez *et al.*，2015；Leiva *et al.*，2015），已经鉴定并验证功能的基因信息可作为参考基因在球囊菌中尝试克隆并进行功能研究。

微卫星标记主要应用于基因连锁与遗传图谱构建（Miyao *et al.*，1996）、谱系及发育研究（Nicolas，1998）、遗传多样性研究（Norri *et al.*，1999）、疾病检测（Hee *et al.*，2001）及品种鉴定（Michael and Guillermo，2003）等。目前已知的球囊菌微卫星标记很少。本研究利用 MISA 软件搜索所有 unigenes，共得到 7 968 个 SSR 位点，随机挑取 20 对引物对两个来源的球囊菌样品中进行 SSR 扩增，有 5 对引物能够扩增出特异性片段。本研究中，SSR 位点扩增产物分离采用的是琼脂糖凝胶电泳，此法被部分国内外研究所采用（辛业芸等，2005；Zhang *et al.*，2013；张鹏飞等，2016），琼脂糖凝胶的分辨率低于 PAGE 胶，可能会导致杂合体不能有效分离。但是，SSRs 位点鉴定可视为对组装的球囊菌参考转录组数据信息的应用，后续研究将通过 PAGE 胶电泳进一步分离 SSRs 位点扩增产物中可能包含的杂合体。传统的微卫星标记开发方法费时耗力、效率低下，基于 RNA seq 数据开发微卫星标记的方法快速高效，有助于大规模开发球囊菌分子标记，推动球囊菌的菌株鉴定、基因图谱构建以及基因定位等研究的深入开展。

白垩病是养蜂生产中的顽疾，尚无有效的防治方法。球囊菌参考基因组和转录组信息的缺失导致严重阻碍该病原的分子研究。本研究成功组装球囊菌的参考转录组，并对其进行了功能与代谢通路注释，将为球囊菌的转录组学研究提供重要的参考信息，为在分子水平深入研究球囊菌基因奠定基础。本课题组前期研究中基于趋势分析的结果表明，意蜂幼虫肠道的球囊菌绝大多数差异表达基因（DEGs）都表现出上升趋势，而中蜂幼虫肠道的球囊菌有相当一部分 DEGs 表现为下降趋势（未发表数据），推测是球囊菌与不同宿主互作的结果，目前正在对球囊菌 DEGs 进行深入分析。

参考文献（References）

Aizen MA，Garibaldi LA，Cunningham SA，Klein AM，2009. How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Ann. Bot.*，103（9）：1579 – 1588.

Alcaíno J，Romero I，Niklitschek M，Sepúlveda D，Rojas MC，Baeza M，Cifuentes V，2014. Functional characterization of the *Xanthophyllomyces dendrorhous* farnesyl pyrophosphate synthase and geranylgeranyl pyrophosphate synthase encoding genes that are involved in the synthesis of isoprenoid precursors. *PLoS ONE*，9（5）：e96626.

Anderson DL，Giacon H，Gibson N，1997. Detection and thermal destruction of the chalkbrood fungus（*Ascosphaera apis*）in honey. *J. Apicult. Res.*，36（3 – 4）：163 – 168.

Anderson DL，Gibson NL，1998. New species and isolates of spore-cyst fungi（Plectomycetes：Ascosphaerales）from Australia. *Aust. Syst. Bot.*，11（1）：53 – 72.

Aronstein KA，Murray KD，2010. Chalkbrood disease in honey bee. *J. Invertebr. Pathol.*，103：S20 – S29.

Baeza M，Fernández-Lobato M，Cifuentes V，2012. Isolation and characterization of extrachromosomal double-stranded RNA elements in *Xanthophyllomyces dendrorhous*. *Methods Mol. Biol.*，898：195 – 205.

Baggio A，Gallina A，Dainese N，Manzinello C，Mutinelli F，Serra G，Colombo R，Carpana E，Sabatini AG，Walner K，Piro R，Sangiorgi E，2005. Gamma radiation：a sanitating treatment of AFB-contaminated beekeeping equipment. *Apiacta*，40：22 – 27.

Bailey L，Ball BV，1991. Honey Bee Pathology. Academic Press Inc.，San Diego，CA，USA. 13.

Chorbiński P，2003. Enzymatic activity of strains of *Ascosphaera apis*. *Medycyna Weterynaryjna*，59（11）：1019 – 1022.

Cornman RS，Bennett AK，Murray KD，Evans JD，Elsik CG，Aronstein K，2012. Transcriptome analysis of the honey bee fungal pathogen，*Ascosphaera apis*：implications for host pathogenesis. *BMC Genomics*，13（7）：285.

Evans JD，Spivak M，2010. Socialized medicine：individual and communal disease barriers in honey bees. *Invertebr. Pathol.*，103：S62 – S72.

Flores JM，Gutierrez I，Espejo R，2005a. The role of pollen in chalkbrood disease in *Apis mellifera*：transmission and predisposing conditions. *Mycologia*，97：1171 – 1176.

Flores JM，Spivak M，Gutierrez I，2005b. Spores of *Ascosphaera apis* contained in wax foundation can infect honeybee brood. *Vet. Microbiol.*，108（1 – 2）：141 – 144.

Gutiérrez MS, Rojas M C, Sepúlveda D, Marcelo B, Víctor C, Jennifer A, 2015. Molecular characterization and functional analysis of cytochrome b5 reductase (CBR) encoding genes from the carotenogenic yeast *Xanthophyllomyces dendrorhous*. *PLoS ONE*, 10(10): e0140424.

Hee SK, Byung LL, Dong KW, Soo IB, Woo HK, 2001. Assessment of markers for the identification of microsatellite instability phenotype in gastric neoplasms. *Cancer Lett.*, 164: 61-68.

Hornitzky M, 2001. Literature review of chalkbrood, a fungal disease of honeybees. A Report by the Rural Industries Research and Development Corporation.

Jensen AB, Aronstein K, Flores JM, Vojvodic S, Palacio MA, Spivak M, 2013. Standard methods for fungal brood disease research. *J. Apicult. Res.*, 52(1): 79-88.

Laf H, 2015. Chalk brood pathogens: a review. *Bee World*, 16(6): 865-869.

Leiva K, Werner N, Sepúlveda D, Salvador B, Marcelo B, Víctor C, Jennifer A, 2015. Identification and functional characterization of the *CYP*51 gene from the yeast *Xanthophyllomyces dendrorhous* that is involved in ergosterol biosynthesis. *BMC Microbiol.*, 15(1): 1-12.

Li JH, Zheng ZY, Chen DF, Liang Q, 2012. Factors influencing *Ascosphaera apis* infection on honeybee larvae and observation on the infection process. *Acta Entomologica Sinica*, 55(7): 790-797. [李江红, 郑志阳, 陈大福, 梁勤, 2012. 影响蜜蜂球囊菌侵染蜜蜂幼虫的因素及侵染过程观察. 昆虫学报, 55(7): 790-797]

Liang Q, Chen DF, 2009. Bee Protection. China Agriculture Press, Beijing. 83-84. [梁勤, 陈大福, 2009. 蜜蜂保护学. 北京: 中国农业出版社. 83-84]

Michael LB, Guillermo O, 2003. Microsatellite loci for paternity analysis in the fathead minnow, *Pimephales promelas* (Teleostei: Cyprinidae). *Mol. Ecol. Notes*, 3(4): 532-534.

Miyao A, Zhong HS, Monna L, Yano M, Yamamoto K, Havukkala I, Minobe Y, Sasaki T, 1996. Characterization and genetic mapping of simple sequence repeats in the rice genome. *DNA Res.*, 3: 233-238.

Mourad AK, Zaghloul OA, Kady EL, Nemat FM, Morsy ME, 2005. A novel approach for the management of the chalkbrood disease infesting honeybee *Apis mellifera* L. (Hymenoptera: Apidae) colonies in Egypt. *Commun. Agric. Appl. Biol. Sci.*, 70(4): 601-611.

Nicolas B, 1998. Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics*, 148: 1893-1906.

Norri AT, Bradley DG, Cunningham EP, 1999. Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture*, 180: 247-264.

Pöggeler S, 2001. Mating-type genes for classical strain improvements of ascomycetes. *Appl. Microbiol. Biotechnol.*, 56: 589-601.

Qin X, Evans JD, Aronstein KA, Murray KD, Weinstock GM, 2006. Genome sequences of the honey bee pathogens *Paenibacillus* larvae, and *Ascosphaera apis*. *Insect Mol. Biol.*, 15(5): 715-718.

Skou JP, 1988. More details in support of the class Ascosphaeromycetes. *Mycotaxon*, 51(1): 191-198.

Spiltoir CF, 1955. Life cycle of *Ascosphaera apis* (*Pericystis apis*). *Am. J. Bot.*, 42(6): 501-508.

Stanley D, Miller J, Tunaz H, 2009. Eicosanoid actions in insect immunity. *Innate Immun.*, 1: 282-290.

Tarpy DR, 2002. Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. *Proc. Biol. Sci.*, 270(1510): 99-103.

Theantana T, Chantawannakul P, 2008. Protease and β-N acetylglucosaminidase of honey bee chalkbrood pathogen *Ascosphaera apis*. *Apicult. Res.*, 47(1): 68-76.

Wang Q, Sun LX, Xiao PX, Liu F, Kang MJ, Xu BH, 2009. Study on technology for indoor artificial feeding of *Apis cerana cerana* larvae. *Shandong Agricultural Science*, (11): 113-116. [王倩, 孙亮先, 肖培新, 刘锋, 康明江, 胥保华, 2009. 室内人工培育中华蜜蜂幼虫技术研究. 山东农业科学, (11): 113-116]

Wood M, 1998. Microbes help bees battle chalkbrood. *Agr. Res.*, 46(8): 16-17.

Xin YY, Zhang Z, Xiong YP, Yuan LP, 2005. Identification and purity test of super hybrid rice with SSR molecular markers. *Chin. J. Rice Sci.*, 19(2): 95-100. [辛业芸, 张展, 熊易平, 袁隆平, 2005. 应用SSR分子标记鉴定超级杂交水稻组合及其纯度. 中国水稻科学, 19(2): 95-100]

Zaghloul OA, MouradAK, El Kady MB, Nemat FM, Morsy ME, 2005. Assessment of losses in honey yield due to the chalkbrood disease, with reference to the determination of its economic injury levels in Egypt. *Commun. Agric. Appl. Biol. Sci.*, 70(4): 703-714.

Zhang L, Yan HF, Wu W, Yu H, Ge XJ, 2013. Comparative transcriptome analysis and marker development of two closely related primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics*, 14(1): 1-13.

Zhang PF, Zhou XR, Pang BP, Tan Y, Chang J, Gao LJ, 2016. High-throughput discovery of microsatellite markers in *Galeruca daurica* (Coleoptera: Chrysomelidae) from a transcriptome database. *Chinese Journal of Applied Entomology*, 53(5): 1058-1064. [张鹏飞, 周晓榕, 庞保平, 谭瑶, 常静, 高利军, 2016. 基于转录组数据高通量发掘沙葱萤叶甲微卫星引物. 应用昆虫学报, 53(5): 1058-1064]

Zheng ZY, Li JH, Liang Q, Chen DF, 2011. *Ascosphaera apis* secretes multiple extracellular enzymes to infect honeybee larvae. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 40(3): 280-284. [郑志阳, 李江红, 梁勤, 陈大福, 2011. 蜜蜂球囊菌分泌多种胞外酶侵染蜜蜂幼虫. 福建农林大学学报(自然科学版), 40(3): 280-284]

(责任编辑: 马丽萍)